

# TarFisDock: a web server for identifying drug targets with docking approach

Honglin Li<sup>1,2</sup>, Zhenting Gao<sup>1</sup>, Ling Kang<sup>2</sup>, Hailei Zhang<sup>2</sup>, Kun Yang<sup>2</sup>, Kunqian Yu<sup>1</sup>, Xiaomin Luo<sup>1</sup>, Weiliang Zhu<sup>1</sup>, Kaixian Chen<sup>1</sup>, Jianhua Shen<sup>1</sup>, Xicheng Wang<sup>2,\*</sup> and Hualiang Jiang<sup>1,3,\*</sup>

<sup>1</sup>Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China, <sup>2</sup>Department of Engineering Mechanics, State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116023, China and <sup>3</sup>School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

Received January 20, 2006; Revised February 13, 2006; Accepted March 10, 2006

## ABSTRACT

TarFisDock is a web-based tool for automating the procedure of searching for small molecule–protein interactions over a large repertoire of protein structures. It offers PDTD (potential drug target database), a target database containing 698 protein structures covering 15 therapeutic areas and a reverse ligand–protein docking program. In contrast to conventional ligand–protein docking, reverse ligand–protein docking aims to seek potential protein targets by screening an appropriate protein database. The input file of this web server is the small molecule to be tested, in standard mol2 format; TarFisDock then searches for possible binding proteins for the given small molecule by use of a docking approach. The ligand–protein interaction energy terms of the program DOCK are adopted for ranking the proteins. To test the reliability of the TarFisDock server, we searched the PDTD for putative binding proteins for vitamin E and 4H-tamoxifen. The top 2 and 10% candidates of vitamin E binding proteins identified by TarFisDock respectively cover 30 and 50% of reported targets verified or implicated by experiments; and 30 and 50% of experimentally confirmed targets for 4H-tamoxifen appear amongst the top 2 and 5% of the TarFisDock predicted candidates, respectively. Therefore, TarFisDock may be a useful tool for target identification, mechanism study of old drugs and probes discovered from natural products.

TarFisDock and PDTD are available at <http://www.dddc.ac.cn/tarfisdock/>.

## INTRODUCTION

Recent advances in the development of tools for docking small molecules to proteins, i.e. virtual screening, has demonstrated the efficiency of this approach for the discovery of potential lead compounds for drug development in the post-genomic era (1–3). Numerous docking programs (4–10) have been used to seek ligands which recognize the 3D structure of a given target obtained by X-ray crystallography, NMR spectroscopy or even by homology modeling [for a review comparing and evaluating docking tools see ref. (11)]. However, identification and validation of druggable targets from amongst thousands of candidate macromolecules is still a challenging task (12,13). A proteomic approach for identification of binding proteins for a given small molecule involves comparison of the protein expression profiles for a given cell or tissue in the presence or absence of the given molecule. This method has not proved very successful in target discovery because it is laborious and time-consuming (14). Thus an efficient computational method for identifying the targets of a small molecule which had been demonstrated experimentally to have an important biological activity would provide a tool of great potential value. An alternative approach that has shown promise in recent years is to use computational methods to find putative binding proteins for a given compound from either genomic or protein databases, and subsequently use experimental procedures to validate the computational result (15–18). One such computational approach,

\*To whom correspondence should be addressed. Tel: +86 21 50807188; Fax: +86 21 50807088; Email: hljiang@mail.shcnc.ac.cn  
Correspondence may also be addressed to Xicheng Wang. Tel: +86 411 84706223; Fax: +86 411 84709390; Email: guixum@dlut.edu.cn

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact [journals.permissions@oxfordjournals.org](mailto:journals.permissions@oxfordjournals.org)

which is the reverse of docking a set of ligands into a given target, is to dock a compound with a known biological activity into the binding sites of all the 3D structures in a given protein database. Protein 'hits' so identified can then serve as potential candidates for experimental validation. Accordingly, this approach is referred to as reverse docking.

Herein, we present a web-based tool Target Fishing Dock (TarFisDock) for seeking potential binding proteins for a given ligand. It makes use of a ligand–protein reverse docking strategy to search out all possible binding proteins for a small molecule from the potential drug target database (PDTD). The small molecule might be a biologically active compound detected in a cell- or animal-based bioassay screen, a natural product or an existing drug whose molecular target(s) is (are) unknown. Thus, TarFisDock may serve as a valuable tool for identifying targets for a novel synthetic compound or for a newly isolated natural product, for a compound with known biological activity, or for an existing drug whose mechanism of action is unknown.

## METHODS

### Construction of the potential drug target database

TarFisDock requires a sufficient number of known protein structures covering a diverse range of drug targets. The target proteins collected in PDTD were selected from the literature (19–22), and from several online databases, such as DrugBank (<http://redpoll.pharmacy.ualberta.ca/drugbank/>) (23), and TTD (<http://bidd.nus.edu.sg/group/cjttd/>) (24). Only proteins with known 3D structures were deposited in PDTD, the Protein Data Bank (PDB) (25) being the major source of their coordinates. PDTD currently consists of 698 entries covering 371 drug targets. These drug targets may be categorized into 15 types, according to their therapeutic areas (20,22), as shown in Table 1. Because TarFisDock does not take into account protein flexibility, PDTD includes redundant entries for proteins known to be flexible. Thus, for example, there are seven entries for HIV-1 (Figure 1).

Water molecules and complexed ligands were removed from the protein structures, after which hydrogen atoms were added, and KOLLMAN charges (26), with the protonation state of the individual residues being taken into account during charge assignment. A mol2 file (Mol2 file (.mol2)

developed by SYBYL, Tripos Inc., St Louis, USA (<http://www.tripos.com/>) is a complete, portable representation of a SYBYL molecule. It is an ASCII file which contains all the information needed to reconstruct a SYBYL molecule.) was then constructed for each protein. The active site of each protein was defined as all residues within 6.5 Å of the ligand bound, and a sphere file for the active site was generated using the SPHGEN program (27). The PDB, mol2 and sphere files for each protein were stored in PDTD.

### Reverse docking procedure using TarFisDock

TarFisDock consists of two parts, a front-end web interface written in both PHP and HTML, with MySQL as database system, and a back-end tool for reverse docking. TarFisDock was developed on the basis of the widely used docking program, DOCK (version 4.0) (5,27). The reverse docking procedure is as follows: (i) TarFisDock either generates a protein target list according to the user's preference (see INPUT) or selects all the protein entries in the PDTD if the user intends to find a new target or targets for an active compound; (ii) TarFisDock docks a given small molecule into the possible binding sites of proteins in the target list, and the interaction energies between the small molecule and the proteins are calculated and recorded; (iii) TarFisDock analyzes the reverse docking result. In general, TarFisDock may output the top 2, 5 or 10% of the ranking list, from which the user may select protein candidates for further biological study. So far, TarFisDock has taken into account the flexibility of the small molecules, but has not yet taken into account protein flexibility. Putative binding proteins are selected by ranking the values of the interaction energy ( $E_{\text{inter}}$ ), which is composed of van der Waals and electrostatic interaction terms (Equation 1),

$$E_{\text{inter}} = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left( \frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332.0 \frac{q_i q_j}{D r_{ij}} \right), \quad 1$$

where each term is a double sum over ligand atoms  $i$  and receptor atoms  $j$ ;  $r_{ij}$  is the distance between atom  $i$  in the ligand and atom  $j$  in the putative receptor protein;  $A_{ij}$  and  $B_{ij}$  are van der Waals repulsion and attraction parameters, respectively;  $a$  and  $b$  are the van der Waals repulsion and attraction exponents, respectively;  $q_i$  and  $q_j$  are point charges on atoms  $i$  and  $j$ ;  $D$  is dielectric function; and 332.0 is the factor that converts the electrostatic energy into kcal/mol. The Amber force field (26) was used for the energy calculation.

**Table 1.** Diseases categories of drug targets in PDTD

(1) Synaptic And Neuroeffector Junctional Sites And Central Nervous System
(2) Inflammation
(3) Renal And Cardiovascular Functions
(4) Gastrointestinal Functions
(5) Uterine Motility
(6) Bacterial Infections
(7) Fungal Infections
(8) Viral Infections
(9) Parasitic Infectious Diseases
(10) Immunomodulation
(11) Blood And Blood-Forming Organs
(12) Neoplastic Diseases
(13) Hormones And Hormone Antagonists
(14) The Vitamins
(15) Undefined

### INPUT, OUTPUT AND OPTIONS

The input file consists of only the test small molecule in standard mol2 format. The 2D structure of a small molecule can be either sketched using ISIS/Draw (ISIS/Draw, MDL Informations Systems, Inc., San Leandro, CA 94557) or ChemDraw (ChemDraw, CambridgeSoft Corporation, 875 Massachusetts Avenue, Cambridge, MA 02139, USA) or taken from such chemical databases as CCD (<http://www.chemnetbase.com/>), ACD (<http://www.mdli.com/>) and SPECS (<http://www.specs.net/>). The user can convert the small molecule from its 2D structures to the 3D structures by using CORINA (28) (<http://www2.chemie.uni-erlangen.>

**PDTD [Potential Drug Target Database]**

Introduction **PDTD** Submit Job Check Job Get Result Forum

Username: Password: [Login](#) [Register](#)

**PDTD**

TarFisDock requires a sufficient number of known protein structures covering a diverse range of drug targets. The target proteins collected in PDTD were selected from the literatures and several online databases, such as DrugBank, and TTD. Only those proteins with known 3D structures were deposited into PDTD. Protein Data Bank (PDB) is the major resource for the 3D coordinates of proteins in PDTD.

PDTD currently consists of 698 entries covering 371 drug targets. According to their therapeutic areas, these drug targets may be categorized into 15 types. Because TarFisDock does not consider the flexibility of proteins, PDTD contains some redundancy for the flexible proteins.

**Search PDTD**

**Fulltext Search**  
  (example: '1a25'; Protein kinase C', [hiv] Disease')  
 or  
**Browse the 15 Categories**  
 Please Select:

The selected category has 22 types of targets

Page 1 (10 rows per page)

PDB_ID	CATEGORY	TARGET_NAME	RELATED DISEASE
1ASX <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV Integrase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1EET <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-1 Reverse Transcriptase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1EP4 <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-2 Reverse Transcriptase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a> , Viral Infection, Unspecified
1EX4 <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV Integrase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1FGL <a href="#">PDB</a> <a href="#">DrugBank</a>	IMMUNOMODULATION	Pepbityl-prolyl cis-trans Isomerase A	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1HNI <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-1 Reverse Transcriptase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1HIV <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	Integrase - Primate lentivirus group Human immunodeficiency virus type 1 (HIV-1)	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1JLA <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-1 Reverse Transcriptase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1LWC <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-1 Reverse Transcriptase	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>
1MTR <a href="#">PDB</a> <a href="#">DrugBank</a>	VIRAL INFECTIONS	HIV-1 Protease	Human Immunodeficiency Virus <a href="#">[hiv] Disease</a>

First Page [Next](#) [Last Page](#)

**Crosslink:**  
 PDB--Protein Data Bank  
 DrugBank--DrugBank database for detailed drug data with comprehensive drug target information.  
 TTD--Therapeutic Target Database

Best View 1024\*768 | Contact  
 Update: 2006-01-20 | Total Visits: 3722  
 Drug Discovery and Design Center (DDDC), SIMM, CAS  
 State Key Laboratory of Structural Analysis for Industrial Equipment, DUT

**Figure 1.** An example of PDTD querying and finding out 22 targets records of '[HIV] DISEASE'.

de/software/corina/free\_struct.html) or other modeling software. The structures can be minimized by means of molecular mechanics, and Gasteiger charges (29) should be assigned to them. Finally, the 3D structure of the small molecule is saved in a mol2 file.

Users can register free of charge for using the TarFisDock server, including access to PDTD. The user must provide his/her email address and username so as to receive the result. After registration, the user can login to the server to upload the mol2 file of the test molecule, customize a target list from PDTD, and submit a job (Figure 2). A job identity number, the 'job\_id', is assigned to each job by the web server, and the number is appended to a job queue in the back-end server. The user may use the job\_id to check the status of his/her job.

The output is delivered in ascending order of energy score (interaction energy). The archive file contains a list of the scores, together with binding models (in mol2 format) of the small molecule tested within the binding sites of the candidate targets. The user can also browse the 'Categories' dropdown menu of PDTD to obtain detailed information for the potential target proteins identified by TarFisDock: the 'PDB\_ID' field contains a hyperlink to the PDB website;

the 'TARGET NAME' field also contains a hyperlink to the DrugBank website (Figures 1 and 2), and any information linking targets to diseases is contained in the 'RELATED DISEASE' field taken from TTD.

## TEST CASES

To test the reliability of the TarFisDock server, we searched for the candidate binding proteins for vitamin E and for 4H-tamoxifen. The results and their comparison with the published experimental data are described below.

### Potential binding proteins for vitamin E

Vitamin E is an antioxidant which is widely used as a dietary supplement (30). It has also been shown to be of therapeutic value in the treatment of a number of diseases, such as cardiovascular disease and some forms of cancer, and to enhance the immune response (31). It is thus likely that vitamin E may interact with multiple target proteins. Indeed, 12 targets for vitamin E have already been reported (16) (Supplementary Table S1). Candidate vitamin E-binding proteins



**TarFisDock [Target Fishing Dock]**

Introduction PDTD Submit Job Check Job Get Result Forum

Username: Password: Login Register

---

**Submit Job**

Your Email:  \* Please provide a valid email address

Ligand File:  \* Please provide a valid file path

Target List: ☒ All Targets

or targets in certain categories

- ☐ BACTERIAL INFECTIONS
- ☐ BLOOD AND BLOOD-FORMING ORGANS
- ☐ FUNGAL INFECTIONS
- ☐ GASTROINTESTINAL FUNCTIONS
- ☐ HORMONES AND HORMONE ANTAGONISTS
- ☐ IMMUNOMODULATION
- ☐ INFLAMMATION
- ☐ NEOPLASTIC DISEASES
- ☐ PARASITIC INFECTIOUS DISEASES
- ☐ RENAL AND CARDIOVASCULAR FUNCTIONS
- ☐ SYNAPTIC AND NEUROEFFECTOR JUNCTIONAL SITE
- ☐ THE VITAMINS
- ☐ UNDEFINED
- ☐ UTERINE MOTILITY
- ☐ VIRAL INFECTIONS

NOTE:

- For Ligand File: ONLY .mol format is supported (example); Gasteiger charge is recommended; Get 3D structure by Swiss Insilico.
- For Target List: All targets contains the targets in all the categories. To specify certain categories of targets, just check the boxes on the left.

---

**TarFisDock [Target Fishing Dock]**

Introduction PDTD Submit Job Check Job Get Result Forum

Username: Password: Login Register

---

**Result of 20051111\_1**

The top 2 % of the targets ranked by energy score in ascending order.

Download (10 MB)

Rank	PDB ID	ENERGY SCORE	TARGET NAME	CATEGORY	RELATED DISEASE
1	1VVR (Protein)	-32.61	Acetylcholinesterase (Protein)	SYNAPTIC AND NEUROEFFECTOR JUNCTIONAL SITES AND CENTRAL NERVOUS SYSTEM	Alzheimer's Disease, Cognitive Deficits
2	1CHT (Protein)	-32.49	Estrogenic 17-Beta Hydroxysteroid Dehydrogenase (Protein)	HORMONES AND HORMONE ANTAGONISTS	Breast Cancer (hormone-sensitive)
3	2NSE (Protein)	-32.07	Nitric Oxide Synthase (Protein)	UNDEFINED	Cardiovascular disease, Colon cancer, Helicobacter infection, Inflammation, Schizophrenia, Sepsis
4	1TCC (Protein)	-31.51	FKBP/tacrolimus binding protein/FK506 binding protein (Protein)	IMMUNOMODULATION	Autoimmune diseases
5	1LDC (Protein)	-30.63	Phenylalanyl-tRNA Synthetase (Protein)	UNDEFINED	
6	2ACS (Protein)	-30.49	Aldose Reductase (Protein)	UNDEFINED	Diabetic Neuropathy, Diabetic Retinopathy, Neuropathic Pain, Noninsulin-dependent Diabetes Mellitus, Diabetic Complications
7	1BAF (Protein)	-30.41	Anti-dinitrophenyl-spin-label monoclonal antibody Fab Fragment (Protein)	UNDEFINED	
8	1THV (Protein)	-30.32	Thermolysin (Protein)	UNDEFINED	
9	1HVR (Protein)	-30.08	HIV Protease (Protein)	VIRAL INFECTIONS	
10	1IMB (Protein)	-30.08	Inositol-1-(or 4)-monophosphatase (Protein)	SYNAPTIC AND NEUROEFFECTOR JUNCTIONAL SITES AND CENTRAL NERVOUS SYSTEM	Bipolar Affective Disorder
11	1HDC (Protein)	-29.89	3-alpha-hydroxysteroid Dehydrogenase (Protein)	GASTROINTESTINAL FUNCTIONS; INFLAMMATION	Androgen-dependent Diseases; Inflammation; Androgen-dependent Diseases; Inflammation
12	1Z96 (Protein)	-29.84	Aldo-Keto Reductase Family 1 Member C3 (Protein)	HORMONES AND HORMONE ANTAGONISTS	Head And Neck Cancer; Cancers
13	1STC (Protein)	-29.84	Camp-dependent Protein Kinase A Type I (Protein)	BACTERIAL INFECTIONS	Hiv-induced T Cell Dysfunction Immune Disease
14	1HSD (Protein)	-29.81	Serine Proteinase alpha-thrombin (Protein)	BLOOD AND BLOOD-FORMING ORGANS	Haemostatic Disorders

\*at least 10 candidates will be output if the targets you selected are more than 10 but the top fraction are less than 10.

Download Result (Top 2 %):  
(This may take up to 30 seconds)

• The docked conformations and the this table in tar ziped archive

Crosslink:  
PDB—Protein Data Bank  
Drugbank—Drugbank database for detailed drug data with comprehensive drug target information.  
TTD—Therapeutic Target Database

Figure 2. An example of the input and output of TarFisDock.

identified using TarFisDock are listed in Supplementary Table S2. The top 2% candidates identified by TarFisDock, ranked by interaction energies, included 4 out of the 12 targets identified experimentally. Three more of these experimentally identified targets were in the top 10% of the proteins ranked by interaction energy. The top 2 and 10% candidates of vitamin E-binding proteins identified by TarFisDock cover 30 and 50%, respectively, of reported targets verified or implicated by experiments. Other targets, such as glutathione *S*-transferase, glutathione synthetase, D-amino acid oxidase, and guanylyl cyclase (it is not available in PDTD), were not identified by TarFisDock (Table 2). The main reason may be that TarFisDock does not take into account protein flexibility. It is of interest that many of the top 10% candidate vitamin E-binding proteins are associated with cancer, cardiovascular diseases, immune function and dementia (Supplementary Table S2).

### Potential binding proteins for 4H-tamoxifen

4H-tamoxifen is used as an adjuvant therapy in the treatment of breast cancer (32). Like vitamin E, it is a multiple target drug. So far, 10 proteins have been identified as interaction targets for 4H-tamoxifen or for its metabolite, tamoxifen (16) (Supplementary Table S1). To test the reliability of our TarFisDock server, we used it to search for candidate binding proteins for 4H-tamoxifen in the PDTD. The target candidates so thus identified are listed in Supplementary Table S3, and those which correspond to proteins identified experimentally are shown in Table 3. Three amongst the top 2% of the candidates are known targets of 4H-tamoxifen, namely dihydrofolate reductase, immunoglobulin and glutathione transferase. The top 5% of the candidates include two additional targets

identified experimentally, i.e. human fibroblast collagenase and 17 $\beta$ -hydroxysteroid dehydrogenase. Of experimentally confirmed targets for 4H-tamoxifen 30 and 50% appear amongst the top 2 and 5% of the TarFisDock predicted candidates, respectively, indicating the reliability of this server tool again.

TarFisDock has been in use for about 9 months, and over 1000 small molecules, including synthetic compounds, existing drugs and natural products, have been screened. Five groups outside the authors' labs have become involved in screening. Experimental evidence has been obtained to confirm that binding proteins identified by TarFisDock for several compounds indeed display binding activity. In one case, that of a binding protein for a natural product, not only was binding verified experimentally, but a complex was obtained whose 3D crystal structure was solved by X-ray crystallography (data not shown). The computing time required depends on the flexibility of the given compound. Thus, TarFisDock may finish the PDTD search within 5–20 h using one CPU of the SGI Origin3800 superserver.

### SUMMARY

In bringing together the target database PDTD and the reverse docking program, TarFisDock server is a convenient tool for identification of potential binding proteins for small molecules such as drugs, lead compounds and natural products. Totally, this web server has already been tested for over 1000 small molecules, the binding proteins for several molecules have been verified by bioassay including crystal structure determination (data not shown). This web server can also

**Table 2.** The protein target candidates of vitamin E identified by TarFisDock

Rank	PDB_ID	Energy score	Target name
1	1VXR	−32.61	Acetylcholinesterase
2	1DHT	−32.49	Estrogenic 17 $\beta$ -hydroxysteroid dehydrogenase
3	2NSE	−32.07	Nitric oxide synthase
6	2ACS	−30.49	Aldose reductase
20	1M9M	−29.28	Nitric oxide synthase
28	1GPN	−28.56	Acetylcholinesterase
49	1DBJ	−27.4	Fab' fragment of monoclonal antibody Db3
50	1ADF	−27.37	Alcohol dehydrogenase
53	4FAB	−27.37	4-4-20 (IgG2A) Fab fragment
62	5P2P	−27.14	Phospholipase A2

**Table 3.** The protein target candidates of 4H-tamoxifen identified by TarFisDock

Rank	PDB_ID	Energy score	Target name
4	1DHF	−36.8	Dihydrofolate reductase
10	1MCR	−35.4	Immunoglobulin $\lambda$ light chain dimer
12	1K3Y	−34.66	Glutathione transferase
13	1DBM	−34.51	Fab' fragment of monoclonal antibody Db3
17	4DFR	−34.07	Dihydrofolate reductase
19	2CGR	−33.68	IgG2B ( $\kappa$ ) Fab fragment
21	4AYK	−33.53	Human fibroblast collagenase
22	4FAB	−33.49	4-4-20 (IgG2A) Fab fragment
25	1DBJ	−33.26	Fab' fragment of monoclonal antibody Db3
27	1DHT	−33.12	17 $\beta$ -Hydroxysteroid dehydrogenase
31	1AYK	−32.9	Human fibroblast collagenase
46	1RA2	−31.75	Dihydrofolate reductase
48	1DIH	−31.71	Microbial dihydrofolate reductase
50	1MCB	−31.66	Immunoglobulin $\lambda$ light chain dimer
70	1BHS	−30.81	17 $\beta$ -Hydroxysteroid dehydrogenase

be used in mapping the regulation genomic network for an existing drug or a drug candidate. In general, one drug molecule may interact with several targets including targets associated with side effect (toxicity). As illustrated by the examples for identifying potential binding proteins of vitamin E and 4H-tamoxifen, TarFisDock provides multiple options for selecting protein targets. These are useful clues for further experimental test in evaluating the efficacy and toxicity of the drug. On the other hand, the targets information produced by TarFisDock is also significant for functional genomic study with the chemical biology paradigm (33). In general, TarFisDock web sever is a convenient tool for 'fishing' the target proteins of small molecules, the user just inputs the structure of querying compound and customizes a target list from PDTT (a list of all the targets is recommended).

However, TarFisDock still has certain limitations. The major one is that the protein entries are not enough for covering all the protein information of disease related genomes. The second one is that TarFisDock has not considered the flexibility of proteins during docking simulation. These two aspects will produce negative false. Another limitation is that the scoring function for reverse docking is not accurate enough, which will produce positive false. To overcome these shortages, we are (i) collecting proteins structures (experimental and modeling structures) as more as possible for enlarging PDTT, (ii) developing new docking program

including protein flexibility, and (iii) establishing accurate scoring function. TarFisDock and PDTT are available at <http://www.dddc.ac.cn/tarfisdock/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

## ACKNOWLEDGEMENTS

The authors thank Prof. Irwin D. Kuntz for providing the source code of DOCK4.0. The Shanghai Supercomputing Center and Computer Network Information Center are acknowledged for allocation of computing time. The authors thank Prof. Israel Silman at Weizmann Institute of Science and the reviewers for critical reading and helpful comments on the manuscript. This study was supported by the Special Fund for the Major State Basic Research Project of China (grants 2002CB512802 and 2002CB512807) from Ministry of Science and Technology of China and the National Natural Science Foundation of China (grant 10572033).

*Conflict of interest statement.* None declared.

## REFERENCES

- Shen, J., Xu, X., Cheng, F., Liu, H., Luo, X., Shen, J., Chen, K., Zhao, W., Shen, X. and Jiang, H. (2003) Virtual screening on natural products for discovering active compounds and target information. *Curr. Med. Chem.*, **10**, 2327–2342.
- Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev. Drug Discov.*, **3**, 935–949.
- Mohan, V., Gibbs, A.C., Cummings, M.D., Jaeger, E.P. and DesJarlais, R.L. (2005) Docking: successes and challenges. *Curr. Pharm. Des.*, **11**, 323–333.
- Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, **261**, 470–489.
- Ewing, T.J., Makino, S., Skillman, A.G. and Kuntz, I.D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, **15**, 411–428.
- Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) Automated docking using a Lamarckian Genetic Algorithm and an empirical binding Free Energy Function. *J. Comput. Chem.*, **19**, 1639–1662.
- Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **267**, 727–748.
- Li, H., Li, C., Gui, C., Luo, X., Chen, K., Shen, J., Wang, X. and Jiang, H. (2004) GASDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorg. Med. Chem. Lett.*, **14**, 4671–4676.
- Dooley, A.J., Shindo, N., Taggart, B., Park, J.G. and Pang, Y.P. (2006) From genome to drug lead: identification of a small-molecule inhibitor of the SARS virus. *Bioorg. Med. Chem. Lett.*, **16**, 830–833.
- Choi, V. (2005) YUCCA: an efficient algorithm for small-molecule docking. *Chem. Biodivers.*, **2**, 1517–1524.
- Kellenberger, E., Rodrigo, J., Muller, P. and Rognan, D. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*, **57**, 225–242.
- Hajduk, P.J., Huth, J.R. and Tse, C. (2005) Predicting protein druggability. *Drug Discov. Today*, **10**, 1675–1682.
- Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nature Rev. Drug Discov.*, **1**, 727–730.
- Huang, C.M., Elmetts, C.A., Tang, D.C., Li, F. and Yusuf, N. (2004) Proteomics reveals that proteins expressed during the early stage of

- Bacillus anthracis infection are potential targets for the development of vaccines and drugs. *Genomics Proteomics Bioinformatics*, **2**, 143–151.
15. Rockey, W.M. and Elcock, A.H. (2005) Rapid computational identification of the targets of protein kinase inhibitors. *J. Med. Chem.*, **48**, 4138–4152.
16. Chen, Y.Z. and Zhi, D.G. (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins*, **43**, 217–226.
17. Chen, Y.Z. and Ung, C.Y. (2001) Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *J. Mol. Graph. Model.*, **20**, 199–218.
18. Paul, N., Kellenberger, E., Bret, G., Muller, P. and Rognan, D. (2004) Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins*, **54**, 671–680.
19. Bonday, Z.Q., Dhanasekaran, S., Rangarajan, P.N. and Padmanaban, G. (2000) Import of host delta-aminolevulinate dehydratase into the malarial parasite: identification of a new drug target. *Nature Med.*, **6**, 898–903.
20. Drews, J. and Ryser, S. (1997) Classic drug targets. *Nat. Biotechnol.*, **15**, 1350.
21. Gibbs, J.B. (2000) Mechanism-based target identification and drug discovery in cancer research. *Science*, **287**, 1969–1973.
22. Hardman, J.G., Limbird, L.E. and Gilman, A.G. (2001) *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. 10 edn. McGraw-Hill, NY.
23. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
24. Chen, X., Ji, Z.L. and Chen, Y.Z. (2002) TTD: Therapeutic Target Database. *Nucleic Acids Res.*, **30**, 412–415.
25. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
26. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
27. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E. (1982) A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, **161**, 269–288.
28. Gasteiger, J., Rudolph, C. and Sadowski, J. (1990) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.*, **3**, 537–547.
29. Gasteiger, J. and Marsili, M. (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, **36**, 3219–3228.
30. Norman, H.A., Butrum, R.R., Feldman, E., Heber, D., Nixon, D., Picciano, M.F., Rivlin, R., Simopoulos, A., Wargovich, M.J., Weisburger, E.K. *et al.* (2003) The role of dietary supplements during cancer therapy. *J. Nutr.*, **133**, 3794S–3799S.
31. Han, S.N. and Meydani, S.N. (1999) Vitamin E and infectious diseases in the aged. *Proc. Nutr. Soc.*, **58**, 697–705.
32. Chen, X., Ung, C.Y. and Chen, Y. (2003) Can an in silico drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? *Nature Prod. Rep.*, **20**, 432–444.
33. Stockwell, B.R. (2004) Exploring biology with small organic molecules. *Nature*, **432**, 846–854.